A PROPOSED MODEL FOR TAILORING CONFIDENTIALITY INFORMATION June 2018

Jennifer Edgar, Bureau of Labor Statistics
Heather Ridolfo, National Agricultural Statistics Service
Robin Kaplan, Bureau of Labor Statistics
Rebecca L. Morrison, National Center for Science and Engineering Statistics
Stephanie Willson, National Center for Health Statistics
Cleo Redline, National Center for Education Statistics
Casey Eggleston, U.S. Census Bureau
Jake Bournazian, Energy Information Administration
Jennifer Hunter Childs, U.S. Census Bureau

ABSTRACT

Using information about the respondent to tailor data collection to the respondent in responsive and adaptive survey designs has been shown to reduce survey error. Allowing interviewers (or web surveys) to tailor question wordings, instructions, or definitions to respondents in conversational interviewing has also been shown to improve data quality. However, tailoring has not been explored in depth in the presentation of confidentiality language prior to beginning the data collection. In this paper, we describe a conceptual model, drawn from the literature and suggested by qualitative evidence. This model illustrates how respondents arrive at their level of assurance about confidentiality and specifies how tailoring confidentiality pledge information may increase the number of respondents who feel assured. The model illustrates that a simple and direct confidentiality statement does assure some respondents, but others need additional information to be assured. However, providing respondents with additional information can also be counterproductive, leading to concerns where there were none or heightening already present concerns. The implication is that to achieve assurance and ultimately gain a completed interview, survey practitioners need to tailor their responses to individual concerns about confidentiality.

INTRODUCTION

Tailoring various aspects of the survey process, such as advance materials, contact strategies, and even survey content, has long been advocated to improve respondent cooperation and

Note: The opinions presented in this paper reflect the views of the authors, and not their respective federal agencies.

response rates (Dillman 1991 and 2000; Dillman, Smyth, and Christian 2014; Groves and McGonagle 2001; Snijkers et al. 2013). In more recent years, tailoring has been implemented in interviewer-administered surveys as well as online self-administered surveys (e.g., Early, Mankoff, and Fienberg 2017). Based on empirical evidence from cognitive interviews, we propose a model to improve the respondent experience by tailoring the process of providing respondents with information about confidentiality. This paper describes a conceptual model, which expands on the work of Groves and McGonagle (2001), of how respondents react to confidentiality pledges, and how survey practitioners and/or interviewers could use that reaction to determine what action to take next to achieve the goal of a completed interview.

Representatives from several federal statistical agencies formed a working group to develop research protocols and test revisions to current confidentiality pledges to respond to requirements arising from the newly passed Cybersecurity Enhancement Act of 2015 (Pub. L. 114-113, Division N, Title II, Subtitle B, Sec. 223). The participating agencies coordinated the cognitive testing of new pledge language to assess respondents' interpretations of, and reactions to, the proposed language. Based on the combined research, the group developed recommendations for language to modify existing confidentiality pledges used by the agencies. For detailed information about study methods and results, see the Statistical Community of Practice Confidentiality Pledge Revision Subcommittee (2016) Final Report. Although the goal of the testing was to ensure that respondents understood any new language and to gain insight on the potential impact on survey response rates, analysis of the data collected revealed inconsistent reactions to the confidentiality pledge language. This paper discusses those results and proposes a model of tailoring the presentation of confidentiality information to address respondent reactions. We also suggest future research to assess the model's effectiveness.

BACKGROUND

CONFIDENTIALITY DISCLOSURE IN SURVEY DATA COLLECTION

As part of the survey process for the U.S. Federal government, and many other survey organizations, respondents are given assurances that their data will be kept confidential. Early work on confidentiality pledges showed that stronger assurances of confidentiality do not necessarily lead to higher response rates (e.g., Singer 1978). While confidentiality may not be a major factor in a person's decision to participate (e.g., Panel on Privacy and Confidentiality as Factors in Survey Response 1979), drawing too much attention to confidentiality assurances or providing elaborate confidentiality assurances can backfire. For example, respondents can become more apprehensive about providing data when they are provided with highly elaborate confidentiality assurances (Singer, Hippler, and Schwarz 1992; Reamer 1979). Additionally, reminding respondents about confidentiality assurances in the middle of a survey can lead to higher nonresponse on some survey questions – including income and demographics (Frey 1986). Conversely, sometimes strong confidentiality assurances can increase response to sensitive questions, such as those about health, income, or drug use (Couper et al. 2008; Joinson, Woodley, and Reips 2007; Tourangeau and Yan 2007). Thus, the literature is somewhat mixed on how respondents react to confidentiality assurances and how this may affect unit and item non-response (Singer, Von Thurn, and Miller 1995).

In light of these findings, Singer and colleagues (1992) put forth the theory that respondents draw meaning from the confidentiality assurances provided. If these assurances are highly elaborate or continually emphasized, respondents are likely to assume that the survey is sensitive or intrusive. Thus, researchers must strike a balance between reassuring respondents that their data are kept confidential while not raising unwarranted concerns. To achieve such a balance, research suggests that respondents prefer to have a simple confidentiality statement including the main points at the beginning of surveys, but like to have the option of reading more information regarding confidentiality while completing the survey (e.g., Landreth, Gerber, and DeMaio 2008).

In addition to survey context and content, respondents' attitudes about confidentiality can affect how they react to confidentiality assurances. The well-documented downward trend in survey response rates (Miller 2017) may be at least partly attributable to the public's increasing concerns about the confidentiality of the data requested (Singer et al. 1992). For example, in the context of government surveys, respondents who had more concerns about the confidentiality of the data and believed that government agencies share information with one another were less likely to participate (Singer 2003; Willimack 2001).

Though some respondents may have negative pre-existing attitudes about data confidentiality, many respondents who participate in surveys have no concerns, and may not have even given much thought to the topic. For example, only about 13 percent of respondents expressed privacy concerns in a survey about health that included sensitive questions (e.g., Bates, Dalhamer, and Singer 2008). This absence of concern may reflect a lack of attention paid to the confidentiality language, because of the repeated exposure to confidentiality statements in general, or the fact that these statements are often dense, difficult to understand, and highly bureaucratic (Gerber 2003; Pascale and Mayer 2004). For instance, Boruch, Dennis, and Cecil (1996) found that many respondents do not recall hearing about confidentiality assurances, nor do they remember the name of the agency that collected the data, even shortly after completing a survey. Similarly, eye-tracking research showed that respondents often just skim over the confidentiality statement because they are more focused on getting started with the survey (Olmstead et al. 2016). Recent evidence suggests that the length of confidentiality statements (and thus the content of the statements) has little effect on unit or item non-response rates to a mail survey (Bucks and Couper 2018), suggesting respondents may not look at these statements closely.

Looking at doorstep refusal conversion, Groves and colleagues (2009) found that privacy and confidentiality concerns were a significant predictor of respondents initially refusing to complete the survey but then ultimately agreeing; “suggesting that if interviewers can successfully address such concerns, they need not lead to nonparticipation” (Groves et al. 2009, p.391). This finding is also supported by a study focused more specifically on confidentiality concerns (Bates, Dahlhamer, and Singer 2008). Much remains to be understood about the best way to design the exchange between interviewers and respondents. However, the existing literature does provide evidence that responding to individual respondents’ concerns, rather than taking a strictly standardized approach, could be an effective way to improve the quality of survey responses and survey participation while fulfilling ethical obligations to appropriately inform respondents.

TAILORING PROCEDURES TO IMPROVE DATA COLLECTION

Rather than using the same data collection strategies for all respondents, responsive and adaptive designs use information about the respondent to tailor data collection to the respondent to improve survey response rates and/or data quality (Groves and Heeringa 2006; Schouten, Calinescu, and Luiten 2013; Wagner et al. 2012; Wagner 2013). In surveys that use responsive design, survey practitioners monitor survey data and paradata during one phase of data collection and use this information to alter strategies in subsequent phases to gain cooperation, and to focus interviewing effort on underrepresented groups with a low response propensity (Groves and Heeringa 2006; Groves et al. 2006). Adaptive survey designs have the same aim, but use information that is known prior to the start of data collection (Peytchev et al. 2010; Shouten, Calinescu, and Luiten 2013; Wagner et al. 2012).

To implement this type of tailoring, Groves and McGonagle (2001) showed that interviewers can be trained to successfully respond to respondent concerns in a way that maximizes survey participation. Training interviewers to adapt their survey request to respondent reactions, and tailoring their responses immediately following evidence of respondent concern, proved to be an effective strategy.

Web surveys provide another opportunity for tailoring. A web survey can be thought of as a dialogue consisting of turns of interaction between the user (respondent) and the survey (or interviewer). Applying Clark’s (1996) collaborative view of comprehension, survey practitioners can design the system to provide options for the respondent based on their understanding and needs. For instance, allowing web survey respondents to get additional information when they want it (e.g., by clicking on hyperlinked definitions), tailoring the system's response to respondents' behavior (e.g., providing pop-up help when they are slow to answer), or by dynamically presenting questions in a certain order based on earlier responses (Early et al. 2017), has been shown to improve the accuracy of respondents' answers (Conrad, Schober, and Coiner 2007).

The process of notifying respondents about confidentiality assurances can be thought of as an ongoing exchange between the interviewer and the respondent, similar to the informed consent process. The goal of this exchange is not only to appropriately notify respondents of any restrictions or exceptions to the confidentiality of their data but also "to give respondents and other subjects meaningful control over information about themselves" (Groves et al. 2009, p. 379) as required by the ethical principle of *Respect for Persons* that governs all research on human subjects, including most survey research (Belmont Report 1979). Although there has been relatively limited empirical research on the best way to improve respondent understanding of confidentiality during the informed consent process, a meta-analysis of the informed consent literature indicates that extended discussion with someone who is knowledgeable about the study may be the most effective method (Flory and Emanuel 2004). Thus, taking a tailored approach to addressing respondents' confidentiality concerns for surveys may also be an effective way to gain assurance.

EVOLVING FEDERAL CONFIDENTIALITY REQUIREMENTS

Many federal statistical agencies in the United States protect data collected from respondents under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA, U.S.C. 3501, 2002). CIPSEA provides "that data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent." This act also allows federal agencies to assure respondents that their data will be kept confidential and only seen by employees or agents of those agencies.

Since 2002, many federal agencies have adopted common language to communicate confidentiality information to respondents in the survey materials, though there is some variation across agencies. In 2015, Congress passed legislation that had the potential to change who had access to respondent data from federal surveys. The Federal Cybersecurity Enhancement Act (CEA) of 2015 (Pub. L. 114-113, Division N, Title II, Subtitle B, Sec. 223), allows the Department of Homeland Security (DHS) to monitor electronic transmissions to and from a federal agency.

Given the potential change in data access, federal statistical agencies had to consider changes to their confidentiality language. A group of six federal statistical agencies worked together to conduct qualitative research to identify language that clearly communicated the intent of the CEA, while not providing so much detail that it created unnecessary concerns. Although this research was conducted by federal agencies for federal surveys, most survey organizations have confidentiality protections for their respondents and communicate them in similar ways, so this paper is relevant to those outside the federal government as well.

METHODS

Based on the CEA, each agency developed wording to add to their current confidentiality pledge group. The language varied slightly by agency, reflecting different sample populations (e.g., household, establishment, farm, or school) and/or legal interpretations of the new legislation. Across all agencies the intended message was consistent - beyond the statistical agency collecting the data, the respondents' survey responses could also be reviewed by another agency as part of cybersecurity monitoring. Some agencies tested multiple versions of the new language, while others focused on only one. All versions tested are shown in the Appendix.

For example, at the Bureau of Labor Statistics, the current CIPSEA language was presented with the new CEA language appended to the end:

> The Bureau of Labor Statistics, its employees, agents, and partner statistical agencies, will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law. In accordance with the Confidential Information Protection and Statistical Efficiency Act of 2002 (Title 5 of Public Law 107-347) and other applicable Federal laws, your responses will not be disclosed in identifiable form without your informed consent. Per the Cybersecurity Enhancement Act of 2015, Federal information systems are protected from malicious activities through cybersecurity screening of transmitted data.

INTERVIEW PROCEDURES

The researchers used cognitive interviews to assess respondents' interpretations of, and reactions to the old and the new confidentiality pledge wording. Each agency employed concurrent and retrospective probing using similar protocols for sessions lasting up to 60 minutes. The interviews focused on the confidentiality language, and did not include any substantive questions from actual agency surveys. Interviews were conducted either on the telephone or in person, with the latter being done in a cognitive laboratory for household respondents or at the respondent's location (e.g., business, school, or farm) for establishment respondents.

Although the exact protocol varied slightly by agency, the following steps were generally followed by all agencies:

1. Respondents were asked general questions about their previous experiences participating in government surveys.
2. Respondents read the agency's current confidentiality pledge language with a version of the new language appended to the end.
3. Respondents answered probes about the confidentiality pledge language, including the following:
    a. Describing their overall reactions to and impressions of the pledge
    b. Whether the pledge would influence their decision to participate in a survey (i.e., their level of assuredness after reading the pledge)
    c. What concerns (if any) they have about the pledge language
    d. What concerns (if any) they have about how the agency would keep their data secure
    e. What concerns (if any) they have about how their data would be used or shared
4. Respondents were then provided with additional information about the confidentiality process. This varied by agency, and could have included an alternate version of the pledge, an explanation of what could be done with their data, and/or how the security monitoring and screening would occur. In all cases however, the additional information[2] provided respondents with new information about how their data would be monitored that was not initially presented.
5. Respondents answered the same probes from step 3 above about the additional information. In addition, interviewers used probes comparing and contrasting the initial pledge (from step 2) with the additional information.
6. Steps 4 and 5 were repeated for the agencies that tested multiple versions of the new confidentiality language, with the order of the versions randomized across respondents.
7. At the end of the interview, respondents gave final comments and elaborated on any concerns they had.

Each agency conducted cognitive interviews with individuals who were representative of their selected survey population; for example NCHS focused on general population and recruited adults 18 years and older with a variety of demographic characteristics, while NCES focused on school surveys and recruited principals, teachers and parents. All agencies aimed to interview a range of respondent types (e.g., race, education, establishment size, industry). Respondents were recruited in a variety of ways, including ads in online bulletin boards and lists of expired

[2] For agencies testing multiple versions, the intention of the authors may have been to present the same information in different ways, but each version actually presented different information. For example, one version referred to federal employees and contractors while another mentioned DHS explicitly. The latter provides more information about who would be doing the cybersecurity monitoring.

survey samples. The sample, number of interviews, and other details of data collection are presented in Table 1.

Table 1. Study Design Summary

Agency	Population	Number of cognitive interviews	Testing Mode	Testing Location
Bureau of Labor Statistics (BLS)	Establishments	22	Phone	
Energy Information Administration (EIA)	Energy companies	25	Phone	
National Agricultural Statistics Service (NASS)	Farm and ranch operators	30	In person	In the field
National Center for Education Statistics (NCES)	School principals, teachers, and parents	24	In person and Phone	In the field and in the lab
National Center for Health Statistics (NCHS)	Adults aged 18 and over	20	In person	In the lab
U.S. Census Bureau	Adults aged 18 and over	30	In person	In the lab

ANALYSIS

To address the study goals, we followed standard qualitative analysis procedures (Miller et al. 2014) to identify themes in respondents' interpretations of, and reactions to, the confidentiality information. This analysis revealed trends and patterns across respondents and agencies, but cannot speak to the magnitude or frequency of those patterns. Detailed results related to respondent interpretation and reactions can be found in the full study report (Statistical Community of Practice Confidentiality Pledge Revision Subcommittee 2016).

In addition to insight into comprehension and interpretation, we identified patterns related to respondent's initial concerns to the confidentiality language and their reactions after receiving additional Information. Using a grounded theory approach (e.g., Glaser and Strauss 2000), we present a model based on these patterns observed in the data and propose research to explore the opportunity to improve respondent experience and, perhaps, response rates by tailoring the confidentially information that is presented.

OBSERVED RESPONDENT REACTIONS TO CONFIDENTIALITY INFORMATION

INITIAL RESPONDENT REACTION

Data from the cognitive interview studies revealed that after interacting with the initial confidentiality pledge, respondents fell into one of two groups: 1) Respondents who had no concerns after reading the pledge (Not Concerned Group) and 2) Respondents who had concerns about confidentiality after reading the pledge (Concerned Group).

Not Concerned Group

The first group of respondents consisted of those who were not concerned after initially reading the confidentiality pledge. In these cases, the initial summary of the pledge was sufficient to elicit respondent assurance because they either had no concerns to begin with, or because they had positive reactions to the information presented in the pledge. The following are examples of respondents who had a positive reaction and felt assured:

> *You get the impression that your privacy and confidentiality is very important in this study.*
>
> *I think my data will be covered. I trust Homeland Security.*
>
> *Makes it sound like a good thing. Makes it sound like you're being protected from malicious activities. People can't just break into the system and get this information about me.*

Concerned Group

The second group of respondents had concerns about the confidentiality of the information they may be asked to provide in a survey. After initially reading the pledge, these respondents were not convinced that their information would be adequately secured.

> *Says the Department of Homeland Security will monitor…just a side note that gives the appearance of great confidence. Except that I keep seeing these places where OMB and other things lose 50 million passwords and that sort of thing. So I am not as personally confident that Department of Homeland Security is going to do such a great job.*

Others expressed confusion about how and when in the process of data collection their data would be monitored.

> *Wait. How are they planning on doing that? The information I'm giving you guys is put into a database or system? When you say it's going to be monitored…but I'm thinking at the end of the day, it's supposed to be done with.*

Others misunderstood the concept of monitoring completely and thought they themselves would be the ones who were monitored.

> *The word 'monitored' kind of makes me not want to take the survey. Was my house chosen because they think I'm a threat to the government?*

OPPORTUNITY TO TAILOR

As noted above, one goal of this research was to determine what, and how much, information to provide respondents about the new CEA legislation. Given our study goal, we presented additional, detailed, information to all respondents regardless of their initial reaction (shown in the Appendix). While this approach of providing additional information to all respondents, regardless of their initial reaction, guaranteed that anyone who might benefit from the additional information received it, it did not lead to uniform respondent reactions. Some respondents who received additional information expressed concerns despite not having any concerns initially, while others continued to find the additional information assuring, and still others who had little to no reactions. Therefore, we propose that survey practitioners take the opportunity to tailor confidentiality information, deciding whether to provide additional details based on whether or not respondents expressed concerns.

FINAL RESPONDENT REACTIONS

The final respondent reactions, either being assured or not assured of confidentiality protection, is of primary interest to all survey practitioners. As demonstrated below, providing more information about confidentiality can lead to different final reactions, and these final reactions could have a positive or negative impact on the probability of response.

In the next section, we walk through the possible paths from respondents' initial reactions to the confidentiality pledge to their final level of assurance after tailoring the confidentiality information.

Concerned Group

Path 1: Concerned → more information → assured

Path 2: Concerned → more information → not assured

Not Concerned Group

Path 3: Not concerned → more information → assured

Path 4: Not concerned → more information → not assured

We will provide qualitative evidence for each path based on the cognitive testing conducted.

Path 1: Concerned → more information → assured

Some respondents, who were concerned after reading the confidentiality pledge, became assured after receiving additional information about the confidentiality of their data. Some of these respondents initially expressed concerns regarding who would be conducting the cybersecurity monitoring and if these individuals would have access to their data.

For example, after initially reading the confidentiality pledge, one respondent expressed the following concerns:

> *[Our identifiable information] could be seen by someone at Homeland Security who's monitoring it…that information is available to an employee of Homeland Security and it could also be used by them…. I would be hesitant and I would want more information [before providing identifiable information].*

After expressing these concerns, this respondent was provided an explanation of the CEA and that additional information sufficiently addressed his concerns:

> *In general, no [concerns]. Two years ago, we did have some information leaked…it could be intentional and unintentional; we could be hacked and I think that's just the risk we take but that is the most efficient way to send information. There's nothing we can do about it but continue to be as proactive as possible.*

Similarly, after first reading the confidentiality statement, another respondent expressed concerns that the statistical agency would be hacked. After receiving additional information indicating that the DHS would be conducting the cybersecurity monitoring, the respondent felt reassured that the data were protected from a hack. Although the additional information resolved concerns on one issue, it created new questions on another issue:

> *If DHS is doing the monitoring, I feel more protected. I wonder if DHS employees are subject to CIPSEA consequences and fines.*

Path 2: Concerned → provide more information → not assured

The path that was more common in our research was those who were concerned after reading the confidentiality pledge and remained not assured about the confidentiality of their data even after receiving additional information. These respondents were often distrustful of the Federal Government and/or the DHS and expressed concerns about who would have access to their information.

For example, one respondent indicated his distrust in the government after initially reading the confidentiality pledge:

> *BS! It's government bureaucracy language. The government will probably form some committee on making this language that will waste taxpayer dollars. They can skew the data or numbers to make it look a certain way.*

The explanation about the new legislation, given after reading the pledge, further fueled his distrust in the government rather than reassuring him.

> *Someone has been hacked and they're covering their butt. There are scams going on, [which] implies that the government was hacked at some point. You don't see these types of disclaimers unless something has already happened.*

Similarly, another respondent was initially concerned about who could see the information he provided on the survey. After receiving additional information, he continued to feel concerned: "Anything can be hacked. I don't care how good your cybersecurity is…"

Path 3: Not concerned → provide more information → assured

Some of these respondents, who did not express concerns initially, remained assured when they were provided with more information. Although in a real survey setting, additional information would probably not be provided to these respondents, the aim of this study was to examine the new language. As such, more information was provided to determine respondent comprehension and reactions. Respondents on this path may legitimately not be concerned, or they may have misguided beliefs about how government agencies share information with each other. For example, after reading the pledge initially one respondent stated, "Nothing here gives me any concerns at all," and after reading an alternative version of the pledge, she continued to feel assured because she assumed that DHS already had access to her information.

> *For me, no [concerns] because the Department of Homeland Security has all my information. Nothing gives me concerns since I am lawful. I just now know that the answers I give will be seen by everyone. That doesn't affect whether I am going to*

> *participate on the survey, I would still do the survey because it's Homeland Security, they probably know everything anyway.*

Similarly, another respondent expressed trust in the Federal Government and the DHS to keep her survey data secure. After reading the pledge initially, she said, "It says that my answers are confidential. Even seeing with DHS, I am being used as a tool for research. I don't feel my safety would be compromised." The respondent continued to feel assured after receiving an explanation of the goals in revising the pledge: "I don't feel the government is out to get anybody. I trust they gather information from me as a teacher to improve education."

Others on this path indicated that the security of their survey data is not something they think about very often. When asked if they had concerns after reading the confidentiality pledge, one respondent stated: "If I thought about it, I might – but [it's] not something I think about – not that big a deal, except if released to certain clients or our competitors." When provided additional information, the respondent felt the changes in the law would make his company's data more secure. "It's a good thing, more protection is better."

Path 4: Not concerned → provide more information → not assured

Some respondents who were not concerned about the security of their data initially changed their stance after being provided additional information, thinking that their data could be compromised. For example, after reading the pledge initially one respondent stated, "If I had any question in my mind about my responses affecting me adversely, [the pledge] might make me feel better." After being provided more information about the law, this respondent became concerned that his data were not as secure as he once thought. "[The pledge] would make me worry that the chances of my data might get hacked – reminds me that this is something that could happen, [whereas] before I hadn't even thought about it." For obvious reasons, this is a path that survey researchers wish to avoid, as it results in creating new concerns that were not present initially.

Similarly, another respondent had no concerns about the agency's ability to secure the survey data. However, after being informed that DHS would be monitoring the data systems, he said, "Sounds like [the statistical agency] lacks the capabilities to protect our data."

Other respondents felt assured when reading that DHS was protecting their survey data but became concerned when reading additional information that stated contractors may be monitoring the data for cybersecurity risks. After receiving the additional information, one respondent said, "Pretty much the DHS will ensure through their security devices that everything will remain confidential." After reading the additional information, this same respondent stated:

> *Okay… Because you just added employees and contractors…even though I know it's someone…it makes me feel more secure to not really know…That is scary, the fact that they have contractors. They could be terrorists for all I know - the contractors. The thing is, this specifies. [And] I am basically telling you I don't like it, that it makes me feel less secure.*

PROPOSED MODEL: TAILORING CONFIDENTIALITY INFORMATION

In light of past research (e.g., Groves and McGonagle 2001), along with the qualitative evidence presented above, we propose a conceptual model (see Figure 1) that describes the process of tailoring the amount of information provided by a confidentiality pledge to respondents with the goal of assuring respondents and increasing the likelihood of response. We suggest that there are three sets of information that precede the initial respondent reaction described above that play a role in the final outcome – these are the survey's and respondent's pre-existing characteristics and the legally required confidentiality pledge itself.

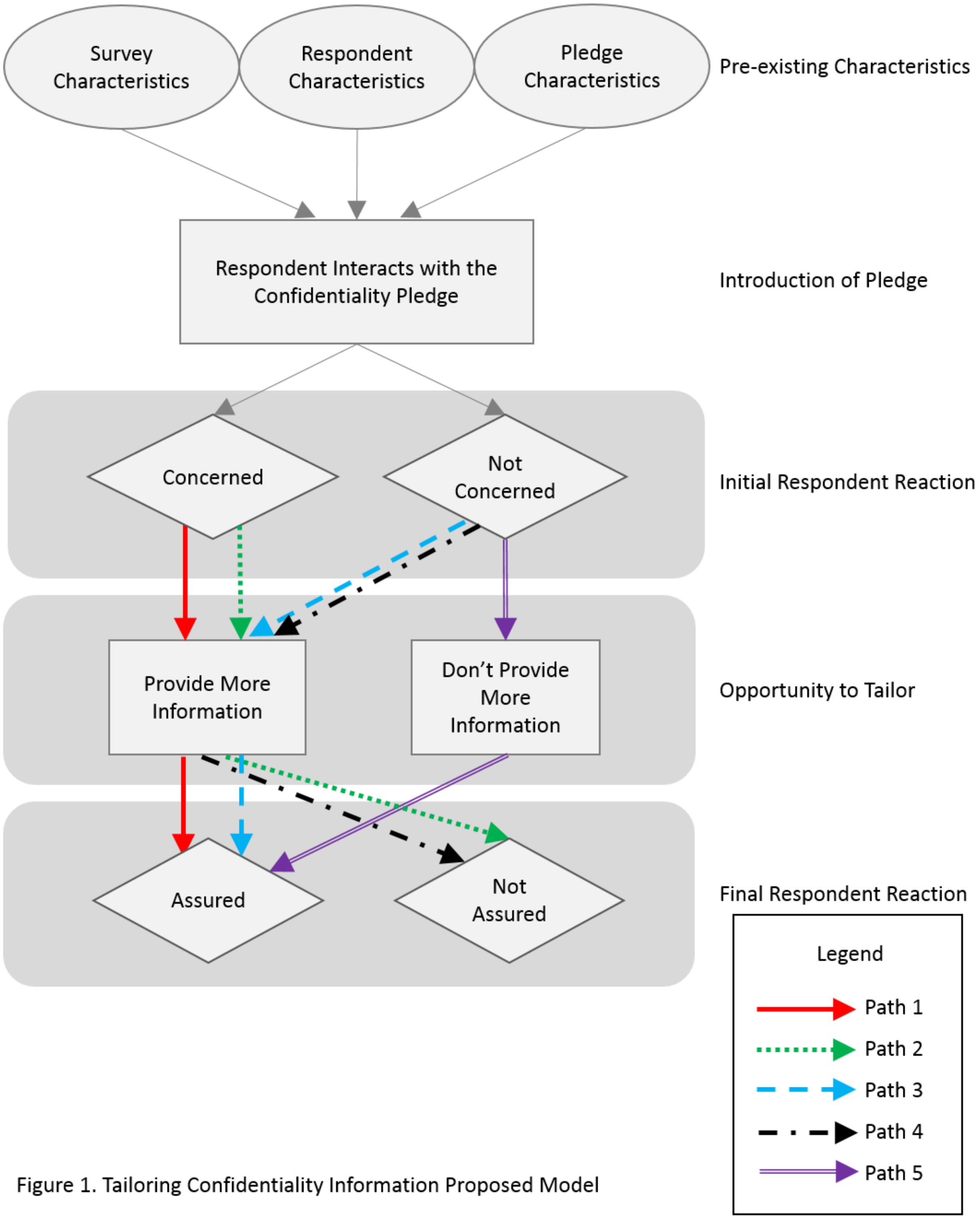
Survey Characteristics
Respondent Characteristics
Pledge Characteristics
Pre-existing Characteristics
Respondent Interacts with the Confidentiality Pledge
Introduction of Pledge
Concerned
Not Concerned
Initial Respondent Reaction
Provide More Information
Don't Provide More Information
Opportunity to Tailor
Assured
Not Assured
Final Respondent Reaction
Legend
Path 1
Path 2
Path 3
Path 4
Path 5

Figure 1. Tailoring Confidentiality Information Proposed Model

CHARACTERISTICS OF THE SURVEY, RESPONDENT AND PLEDGE

This model posits that several characteristics exist prior to the survey solicitation that impact respondents' interactions with confidentiality pledges. The first – characteristics of the survey – is well established in the literature. Features such as survey sponsor and topic have been shown to impact how respondents react to confidentiality information (Dillman, Smyth, and Christian 2014). For example, surveys with sensitive topics require more extensive assurances of confidentiality (Singer et al. 1995), while surveys with non-sensitive topics generally require less (Singer et al. 1992).

The second – respondent characteristics – is equally important. Respondents' past experiences, attitudes and opinions about surveys, specific survey sponsors, data security or confidentiality have a direct impact on how they react to confidentiality information (Joinson et al. 2008; Couper et al. 2008). For example, respondents who have had their identity stolen may be more likely to pay more attention to and have more concerns about how their information will be kept confidential than those who have not had that experience. On the other hand, respondents who have had positive experiences with surveys or who appreciate the value of the data collected may have fewer confidentiality concerns (Couper, Singer, and Kulka 1998).

The final characteristic is the pledge content itself. There is not extensive research on this topic, but as noted above, elaborate confidentiality pledges and multiple reminders of confidentiality protection during the survey process can impact respondents' initial reactions and subsequent decisions to participate (Singer, Hippler, and Schwarz 1992; Reamer 1979). The exact language used in a confidentiality pledge is often determined by statisticians, attorneys, or institutional review boards. As a result, pledges frequently include technical and legal jargon or references to laws that may or may not be meaningful to respondents. The amount of detail, understandability, word choice, and other features of the pledge may also affect respondents' reactions. Whether it is included in the advance letter, on the online survey introductory screen, or read to the respondent by an interviewer, respondents are typically exposed to a standardized confidentiality pledge as a routine part of the survey process. The characteristics of the pledge may influence respondents' interactions with it, that is, how much of the pledge is read or how attentive they are to details of the pledge. This in turn may shape their initial reactions to the pledge wording, specifically whether or not they are concerned about the confidentiality of their survey data.

As described above, we observed four paths after all our respondents were provided with additional information. However, our model posits that if a respondent is not concerned after initially reading the confidentiality pledge, the respondent does not need any more information to be assured and should instead start the survey immediately.[3] This constitutes a fifth path that was not included in our testing but is likely to occur in the field:

PATH 5: NOT CONCERNED → DO NOT PROVIDE MORE INFORMATION→ ASSURED

Given the ultimate goal of achieving respondent assurance with the confidentiality pledge, we suggest that a tailored approach be implemented once the respondent's reaction is gauged. Rather than give all respondents more confidentiality information – which may unwittingly create concerns among some respondents who initially did not have any (path 3 above) – targeting additional information only to respondents with concerns is likely to be a more effective approach. The same targeting could also be applied for subsequent respondent reactions. If respondents are not assured after receiving more information, the interviewer would identify the nature of their concerns and offer additional information to try to assure them, ideally by speaking to their expressed concerns. Although there will always be some who remain concerned, we argue that targeting appropriate additional information to those respondents is likely to reduce the size of that group.

FUTURE RESEARCH

While we have provided preliminary support of this model, we recognize that research is needed to both develop implementation strategies as well as evaluate the effectiveness of the tailoring approach. In this section we outline a research plan that could be followed.

First, qualitative research is needed to develop both the initial confidentiality language and the additional information to be provided when needed. Using the framework of Groves and McGonagle's (2001) theory-guided interviewer training protocol of survey participation, researchers should conduct exploratory research with both respondents and interviewers to develop an inventory of types of concerns, as well as additional information that can be provided to assuage them. The results from this research could be used to develop interviewer training to ensure interviewers are able to accurately identify respondent concerns through active listening and provide effective, tailored, additional information to address the concern. Ensuring interviewers are skilled at identifying and classifying the concern is key to the success of this approach, but we recognize that there are situations where the tailored approach will not be feasible, because some respondents simply do not provide enough information about the reason for refusal. Additionally, as with all forms of adaptive design, or non-standardized

[3] We acknowledge the existence of another theoretical path (concerned → do not provide more information → not assured), but given the survey practitioner's goal of obtaining respondent cooperation, we believe this is not a path that would be followed.

interviewing, there is the potential for new types of error to be introduced. Allowing interviewers to decide what additional information to provide could affect respondents in unknown ways (e.g., respondents cooperate but refuse to provide sensitive information).

To test the effectiveness of this tailored approach, a field experiment is needed. By implementing the tailored approach for half the data collection and capturing information about respondents' initial concerns and their subsequent reactions and response decision, researchers could determine if the tailored approach to providing confidentiality information is effective. Ideally, multiple surveys would test this approach, allowing for factors such as content, sponsor, topics and sensitivity to be analyzed as covariates.

Although tailoring is well suited for interviewer administered surveys, we see potential for use in online self-administered surveys as well. As mentioned, online surveys have become more dynamic in recent years, mimicking the experience of having an interviewer-respondent exchange. Following this logic, a dynamic online survey could display a pop-up message if a respondent spends a long time on the page that presents the confidentiality language. The pop-up could ask if respondents have any questions about the page. If respondents indicate they have a question about confidentiality, the survey could ask about the nature of the concern (using the categories identified from the exploratory research described above). Then a corresponding tailored message with additional information addressing the particular concern could be displayed, and respondents would have the opportunity to indicate the extent to which the additional information assured them. Researchers could then collect paradata to determine whether respondents proceeded with the survey or dropped out. Using a similar approach, Lewis, Gorsak, and Yount (2018) demonstrated that presenting a tailored appeal to respondents considering opting out of an online survey due to confidentiality concerns was an effective refusal conversion strategy. In addition, other variables such as survey topic, respondent characteristics, and pledge characteristics could be assessed as covariates, as well as any paradata the online instrument collected that would inform respondent experience with the survey.

There are also adaptations of this model that are worth exploring. For example, researchers could consider tailoring initial confidentiality information to different respondent groups. For instance, if certain respondents tend to have specific types of confidentiality concerns (e.g., immigrant households who have concerns about their data being shared across federal agencies), the pre-scripted confidentiality language could address that topic upfront rather than in the additional information, allowing for application of this model even in paper forms.

CONCLUSIONS AND RECOMMENDATIONS

Although confidentiality concerns may not be a primary factor driving respondents' response decision, it is an aspect of the survey process that should be explored as survey practitioners continue efforts to address declining response rates. The language used to inform respondents that their information will be kept confidential is likely one of the least researched features of survey design. This language, often written to address all aspects of the relevant laws and policies, is added to carefully crafted advance materials and survey forms without much consideration of the impact on the respondent. We propose that there is room for improvement in this area. Just as we tailor letters, survey forms and data collection procedures, we can also tailor confidentiality information. Rather than giving all respondents the same information, we suggest identifying the core components of the confidentiality information to provide to everyone as a starting point, and then providing additional information for those respondents who indicate initial concern. By tailoring the process to the individual respondent, rather than to the agency or the survey content as suggested by Singer, Von Thurn, and Miller (1995), we believe there are potential improvements to be realized.

The model proposed illustrates the process of how respondents react to confidentiality pledges and specifies how tailoring the process of providing confidentiality information may increase the number of respondents who feel assured. The proposed model agrees with Singer's (2003) proposition that 'less can be more.' For some respondents, a simple and direct confidentiality statement is sufficient, and more information may create concerns that would not otherwise exist. However, for other respondents, additional information is necessary to provide assurance. Our model builds on this strategy by suggesting a flexible protocol. Using a tailored approach, an interviewer (or dynamic instrument) would assess respondents' initial reactions to the pledge and determine whether to offer more information in an attempt to assuage concerns.

Though the specific additional information provided varied across the participating agencies, what we find remarkable is that all agencies observed some respondents switching from assured to no-longer-assured. While we do not have experimental evidence to support the model, we do have qualitative evidence to suggest that the common approach of giving everyone the same, additional detailed, information may not be ideal. The presentation of more information needs to be carefully considered as it may actually create concerns among respondents who initially were not concerned.

We propose that there is an opportunity to improve respondents' experience by tailoring presentation of confidentiality information. This seems to be an understudied area that, with minimal effort, could also lead to improvements in a person's understanding of the confidentiality pledge and his or her willingness to participate in a survey. Future research should test this model, taking into account the pre-existing conditions (respondent, survey and pledge characteristics) and the effects of taking the opportunity to tailor (providing additional information or not). We leave it to future researchers to assess respondents' concerns about confidentiality and to determine exactly what information will successfully counter those specific concerns, using plain and simple language that avoids jargon and legalese.

REFERENCES

Bates, Nancy, James Dahlhamer, and Eleanor Singer. 2008. "Privacy Concerns, Too Busy, Or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse." *Journal of Official Statistics* 24(4): 591- 612.

The Belmont Report. 1979. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

Boruch, Robert F., Michael Dennis, and Joe S. Cecil, 1996. "Fifty Years of Empirical Research on Privacy and Confidentiality in Research Settings." *Research Ethics: A Psychological Approach.* Lincoln and London: University of Nebraska Press. 129-73.

Clark, Herbert H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Confidential Information Protection and Statistical Efficiency Act. 2002. 44 U.S.C. 3501 (note).

Conrad, Frederick G., Michael F. Schober, and Tania Coiner. 2007. "Bringing Features of Human Dialogue to Web Surveys." *Applied Cognitive Psychology* 21: 165-88.

Conrad, Frederick G. and Michael F. Schober. 2000. "Clarifying Question Meaning in a Household Telephone Survey." *Public Opinion Quarterly* 64: 1-28.

Couper, Mick P., Eleanor Singer, Frederick G. Conrad, and Robert M. Groves. 2008. "Risk of Disclosure, Perceptions of Risk, and Concerns about Privacy and Confidentiality as Factors in Survey Participation." *Journal of Official Statistics* 24(2): 255-75.

Couper, Mick P., Eleanor Singer, and Richard A. Kulka. 1998. "Participation in the 1990 Decennial Census: Politics, Privacy, Pressures." *American Politics Quarterly* 26: 59-80.

Dillman, Don A. 1991. "The Design and Administration of Mail Surveys." *Annual Review of Sociology* 17(1): 225-49.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. New York: John Wiley & Sons, Inc.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons, Inc.

Early, Kirstin, Jennifer Mankoff, and Stephen E. Fienberg. 2017. "Dynamic question ordering in online surveys." *Journal of Official Statistics* 33(3): 625-657.

Federal Cybersecurity Enhancement Act of 2015. 6 U.S.C. 1521 et. seq.

Federal Trade Commission. "*How to Keep Your Personal Information Secure*." Accessed January 30, 2018. https://www.consumer.ftc.gov/articles/0272-how-keep-your-personal-information-secure

Flory, James, and Ezekiel Emanuel. 2004. "Interventions To Improve Research Participants' Understanding in Informed Consent for Research: A Systematic Review." *JAMA* 292(13): 1593-601.

Frey, James H. 1986. "An Experiment with a Confidentiality Reminder in a Telephone Survey." *Public Opinion Quarterly* 50: 267-69.

Gerber, Eleanor. 2003. "Respondents Understanding of Confidentiality Language." *Annual Conference of the American Association for Public Opinion Research.* Nashville, Tennessee. American Association for Public Opinion Research.

Glaser, Barney G., and Anselm L. Strauss. 2000. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago, IL: Aldine Publishing.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Groves, Robert M., and Steven G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169: 439–57.

Groves, Robert M., and Katherine A. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17: 249-265.

Joinson, Adam N., Alan Woodley, and Ulf-Dietrich Reips. 2007. "Personalization, Authentication and Self-Disclosure in Self-Administered Internet Surveys." *Computers in Human Behavior* 23: 275-85.

Joinson, Adam N., Carina Paine, Tom Buchanan, and Ulf-Dietrich Reips. 2008. "Measuring Self-Disclosure Online: Blurring and Non-Response to Sensitive Items in Web-Based Surveys." *Computers in Human Behavior* 24: 2158-171.

Landreth, Ashley, Eleanor Gerber, and Theresa DeMaio. 2008. "*Report of Cognitive Testing of Privacy and Confidentiality-Related Statements in Respondent Materials for the 2010 Decennial: Results from Cognitive Interview Pretesting with Volunteer Respondents*." US Census Bureau. Statistical Research Division Report Series (Survey Methodology# 2008-4). On-line, available: www.census.gov/srd/papers/pdf/rsm2008-04.

Lewis, Taylor, Mark Gorsak, and Naomi Yount. 2018. "An Automated Refusal Conversion Strategy for Web Surveys." Proceedings from the Federal Committee on Statistical Methodology Research and Policy Conference.

Miller, Kristen, Stephanie Willson, Valerie Chepp, and Jose-Luis Padilla. 2014. *Cognitive Interviewing Methodology: A Sociological Approach for Survey Question Evaluation.* Hoboken, NJ: John Wiley & Sons, Inc.

Miller, Peter V. 2017. "Is There a Future for Surveys?" *Public Opinion Quarterly* 81: 205-12.

Olmsted-Hawala, Erica, Lin Wang, Diane K. Willimack, Emily Stack, and Sabin Lakhe. 2016. "A Pilot Investigation of the Association between Eye-Tracking Patterns and Self-reported Reading Behavior." In *International Conference on Universal Access in Human-Computer Interaction*, 442-53. Toronto, Ontario, Canada. Springer International Publishing.

Peytchev, Andy, Sarah Riley, Jeff Rosen, Joe Murphy, and Mark Lindblad. 2010. "Reduction of Nonresponse Bias through Case Prioritization." *Survey Research Methods* 4: 21-29.

Reamer, Frederic G. 1979. "Protecting Research Subjects and Unintended Consequences: The Effect of Guarantees of Confidentiality." *Public Opinion Quarterly* 43: 497-506.

Schober, Michael F. and Frederick G. Conrad. 1997. “Does Conversational Interviewing Reduce Survey Measurement Error?” *Public Opinion Quarterly* 61: 576-602*.*

Schouten, Barry, Melania Calinescu, and Annemieke Luiten. 2013. "Optimizing Quality of Response through Adaptive Survey Designs." *Survey Methodology* 39: 29-58.

Singer, Eleanor. 1978. "Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys." *American Sociological Review* 42(2): 144-62.

Singer, Eleanor, Dawn R. Von Thurn, and Esther Miller. 1995. "Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature." *Public Opinion Quarterly* 59(1): 66-77.

Singer, Eleanor, Hans-Juergen Hippler, and Norbert Schwarz. 1992. "Confidentiality Assurances in Surveys: Reassurance or Threat?" *International Journal of Public Opinion Research* 4: 256-68.

Singer, Eleanor. 2003. "Exploring the Meaning of Consent: Participation in Research and Beliefs about Risks and Benefits." *Journal of Official Statistics* 19(3): 273-85.

Snijkers, Ger, Gustav Haraldsen, Jacqui Jones, and Diane Willimack. 2013. *Designing and Conducting Business Surveys*. New York: John Wiley & Sons.

Statistical Community of Practice Confidentiality Pledge Revision Subcommittee (2016), “Final Report.” In the authors’ possession.

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5): 859-83.

Wagner, James.2013. “Adaptive Contact Strategies in Telephone and Face-To-Face Surveys.” *Survey Research Methodology* 7(1): 45-55.

Wagner, James, Brady T. West, Nicole Kirgis, James M. Lepkowski, William G. Axinn, and Shonda K. Ndiaye. 2012. “Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection*.” Journal of Official Statistics* 28(4): 477-99.

Willimack, Diane. K. 2001. “Businesses’ Perceptions of Confidentiality and Their Attitudes towards Data Sharing Among Federal Statistical Agencies.” Proceedings from the Federal Committee on Statistical Methodology.

APPENDIX

Six federal statistical agencies were involved in cognitively testing new confidentiality pledge wording. Since each agency conducted their surveys under different legislative authorities, and had input from their respective legal, ethical, and technical staff, there were different versions tested. Across agencies however, the new language reflecting the Federal Cybersecurity Enhancement Act of 2015 (CEA) was added to their current confidentiality pledge information. The new language tested at each agency are listed below, along with the population of interest.

BUREAU OF LABOR STATISTICS (ESTABLISHMENT RESPONDENTS):

Per the Cybersecurity Enhancement Act of 2015, Federal information systems are protected from malicious activities through cybersecurity screening of transmitted data.

ADDITIONAL INFORMATION:

What we're trying to communicate is that the Department of Homeland Security will be monitoring the electronic systems where survey data is transmitted and stored. They won't be looking at individual survey answers, instead they will be monitoring the systems to look for viruses, malware and other threats.

ENERGY INFORMATION ADMINISTRATION (ENERGY COMPANIES):

Version 1: EIA information systems are further protected by Federal employees and contractors through security monitoring of the systems that transmit your data.

Version 2: Your data are further protected by Department of Homeland Security cybersecurity employees and contractors through security monitoring of the systems that transmit your data.

ADDITIONAL INFORMATION:

All respondents received both versions of the revised pledge. The order in which the respondents read the revised pledge was reversed for half of the interviews.

In addition, all respondents were provided the following information after reading both revised pledges: What we're trying to communicate is that the Department of Homeland Security will be monitoring the electronic systems where survey data is transmitted and stored. They won't be looking at individual survey answers, instead they will be monitoring the systems to look for viruses, malware and other threats.

NATIONAL AGRICULTURAL STATISTICS SERVICE (FARM AND RANCH OPERATORS):

Version A: Your data are further protected by Department of Homeland Security employees and contractors through cybersecurity monitoring of the systems that transmit your data.

Version B: Per the Cybersecurity Enhancement Act of 2015, Federal information systems are protected from malicious activities through cybersecurity screening of transmitted data.

ADDITIONAL INFORMATION:

All respondents received both versions of the revised pledge. The order in which the respondents read the revised pledge was randomized. In addition, all respondents were provided two pieces of additional information, always in the same order:

In the event of a cybersecurity incident, and pursuant to any required legal process, information from these sources may be used to help identify and mitigate the incident.

What we're trying to communicate is that the Department of Homeland Security will be monitoring the electronic systems where survey data is transmitted and stored. They won't be looking at individual survey answers, instead monitoring the systems to look for viruses, mal-wear and other threats.

NATIONAL CENTER FOR EDUCATION STATISTICS (SCHOOL PRINCIPALS, TEACHERS, AND PARENTS):

Revised Statement A: Electronic transmission of your information will be monitored by Homeland Security in accordance with the Cybersecurity Enhancement act of 2015.

Revised Statement B: NCES systems are further protected by Federal employees and contractors through security monitoring of systems that transmit your data.

ADDITIONAL INFORMATION:

All respondents received both versions of the revised pledge. The order in which the respondents read the revised pledge was randomized.

In addition, all respondents were provided the following information after reading both revised pledges: What NCES is trying to communicate is that the Department of Homeland Security will be monitoring the electronic systems where survey data are transmitted and stored. The

Department of Homeland Security will be monitoring the systems for viruses, malware, and other threats.

NATIONAL CENTER FOR HEALTH STATISTICS (ADULTS, AGE 18+):

Version 1: Electronic transmission of your information will be monitored in accordance with the Cybersecurity Enhancement Act of 2015.

Version 2: Except to comply with the Cybersecurity Enhancement Act of 2015 which monitors information sent to federal agencies for cybersecurity threats, the information will be held confidential.

Version 3: Per the Cybersecurity Enhancement Act of 2015, Federal information systems are protected from malicious activities through cybersecurity screening of transmitted data.

ADDITIONAL INFORMATION:

All respondents received all three versions of the revised pledge. The order in which the respondents read the revised pledge was randomized.

U.S. CENSUS BUREAU (ADULTS, AGE 18+):

DHS Version: Further, your data are protected from cybersecurity risks by the Department of Homeland Security through security monitoring of the systems that transmit your data.

Cybersecurity Version: Further, your data are protected from cybersecurity risks by security monitoring of the Census Bureau information systems.

Web Version: Any information you enter into this system is confidential and may be used by the Census Bureau for statistical purposes, as well as for other uses, such as improving the efficiency of our programs...Use of this system indicates your consent to us collecting, monitoring, recording, and using the information that you provide for any lawful government purpose.

ADDITIONAL INFORMATION

All respondents received all three versions of the revised pledge. The order in which the respondents read the revised pledge was randomized.